Balancing our perceptions of the efficacy of success-based feedback manipulations on motor learning

Ong, N. T., & Hodges, N. J. (2018). Balancing our perceptions of the efficacy of success-based feedback manipulations on motor learning. *Journal of motor behavior*, *50*(6), 614-630.
https://www.tandfonline.com/doi/full/10.1080/00222895.2017.1383227

Nicole T. Ong

and

Nicola J. Hodges[1]

University of British Columbia

Keywords: social-comparison, self-efficacy, perceived competency, motivation, errorless learning

[1]Primary contact

Address: 210-6081 University Blvd, Vancouver, BC V6T 1Z1, Canada

Telephone: +1 (604) 822-5895

Fax: +1 (604) 822-6842

E-mail: nicola.hodges@ubc.ca

Abstract

We report two experiments evaluating the impact of success-related feedback on learning of a balance task. In Exp. 1, we studied the influence of lax and conservative outcome feedback, as well as large vs. small improvements in outcome feedback on balance learning. Despite impacts on competency, there were no between-group differences in actual performance or learning. Because no comparative information was provided in Exp. 1, we tested four further groups that either did or did not receive positive or negative comparative feedback (Exp. 2). Although the manipulations influenced competency and arousal, again, there was no impact on balance outcomes. These data cast doubt on the assertions made in the OPTIMAL theory that perceptions of success are moderators of motor learning.

Balancing our perceptions of the efficacy of success-based feedback manipulations on motor learning

In the behavioural study of motor learning, much of the established research has been focused on cognitive processes that impact the effectiveness of practice. Recently, this focus has shifted toward a social-cognitive and affective perspective. Spearheading the study of social-cognitive influences on motor skill acquisition, Wulf and colleagues have conducted a number of studies which seem to point towards an important role of competency and success perceptions on motor learning[1] (e.g. Chiviacowsky & Wulf, 2007; Lewthwaite & Wulf, 2010b, 2012). This has culminated in the proposal of the OPTIMAL theory of motor learning (Wulf & Lewthwaite, 2016). In addition to short-term (immediate) effects of success perceptions on performance, potentially associated with increased effort during practice, the authors have shown retention effects over longer time periods, with or without performance effects, suggestive of enhanced memory consolidation associated with feelings of success.

Several creative methods have been used to enhance perceptions of success in motor learning, through selective provision of feedback on trials with better (or worse) outcomes, through interventions involving manipulations to the target goal, or through the provision of false comparative (normative) feedback (e.g., Badami, VaezMousavi, Wulf & Namazizadeh, 2012; Chiviacowsky & Wulf, 2007; Chiviacowsky, Wulf, Wally & Borges 2009; Saemi, Porter, Ghotbi-Varzaneh, Zarghami & Maleki, 2012). One of the first motor learning studies to show benefits from false comparative feedback was conducted using a stabiliometer, balance task (Lewthwaite & Wulf, 2010a). Here, feedback that participants were doing better than others positively impacted performance and learning (compared to ‘negative’ feedback and no comparative feedback groups). The authors concluded that both performance and learning were aided by enhanced perceptions of success and motivation. However, no measures of success perceptions were taken and there was no indication as to how this feedback immediately impacted balance performance, which may reflect existing between group differences.

Motor learning has also been enhanced through target/goal-criteria settings that render greater success experiences for a group engaged in an easier, rather than more difficult version of a task (e.g., larger targets or more lax accuracy requirements). In one such study, novices practised golf-putting to either large or small targets, resulting in success on 22% versus 8% of trials respectively (Palmer, Chiviacowsky & Wulf, 2016). Not only did the easier target group make more successful ‘hits’ during practice, it was more accurate in delayed retention when putting to the small target. The authors concluded that performance and learning were positively influenced by expectancy to succeed in the task and that this might have led to a more ‘automatic’ mode of control that supported learning. However, there was no evidence that the learning effects were driven by enhanced feelings of success, rather than perhaps different control strategies associated with task accuracy demands. Without data on outcome or movement variability, it is also not possible to rule out the interpretation that learning was compromised due to decreased stability in performance, as a result of more trial-to-trial corrections associated with attainment of a more stringent goal.

Practice protocols or techniques that encourage fewer errors, or less knowledge of errors, are thought to be good for long-term learning, because they encourage a more “automatic” or proceduralized mode of control (e.g., Maxwell, Masters & Eves, 2000; Maxwell, Masters, Kerr & Weedon, 2001; Poolton, Masters & Maxwell, 2005). This type of learning, which has been studied in reference to implicit motor learning, is particularly robust under pressure or working memory demands (such as dual-tasks). Negative feedback or unsuccessful performances are believed to promote conscious control of movement and accumulation of verbalizable, explicit knowledge. This knowledge is then prone for “reinvestment” under test or pressure sensitive conditions and leads to diminished learning and stability over the long-term (e.g., Gray, 2011; Masters, 1992; Masters & Poolton, 2012; Maxwell & Masters, 2009). Therefore, any negative effects associated with error manipulations may be a function of knowledge build-up, rather than success perceptions. Without assessment of performance under dual-task or pressure conditions, or measurement of verbalizable knowledge, it has not been possible to rule out alternative explanations for these effects.

Interestingly, long-term effects associated with manipulations to success perceptions have emerged in the absence of short-term effects (Ávila, Chiviacowsky, Wulf & Lewthwaite, 2012; Badami et al., 2012; Chiviacowsky & Wulf, 2007; Trempe, Sabourin & Proteau, 2012; Wulf, Chiviacowsky & Lewthwaite, 2010). For motor skill learning, the post-practice retention period has been shown to be a crucial time period, allowing motor skills to be encoded into stable memory forms, facilitating successful retrieval (Brashers-Krug, Shadmehr & Bizzi, 1996; McGaugh, 2000; Robertson, Press & Pascual-Leone, 2005; Shadmehr & Holcomb, 1997). Both reward and positive affect have been associated with improved memory consolidation, through neural modulations of the dopaminergic (e.g., Holroyd & Coles, 2002; Hosp, Pekanovic, Rioult-Pedotti & Luft, 2011; Murayama & Kitagami, 2014; Wise, 2004) and hormonal systems (Cahill & Alkire, 2003; McGaugh & Roozendaal, 2002 for a review). Therefore, it has been argued that perceptions of success complement instances of reward and aid retention through processes of memory consolidation, when people believe they have been successful.

Data consistent with memory consolidation ideas was provided in a visuo-motor adaptation aiming task, via manipulations to target size (i.e., goal criteria). Here an easier goal group was more accurate in retention compared to a difficult goal group, even though it did not differ in practice and retention was tested under difficult goal conditions (Trempe et al., 2012). In such tasks, greater precision in aiming and large magnitude errors, have been associated with a more strategic type of learning (and different neural structures) that is less robust over time than aiming to targets with reduced constraints and smaller magnitude errors (e.g., Criscimagna-Hemminger, Bastian & Shadmehr, 2010). Hence, although goal-success and perceptions of competency may offer one reason for the goal-related effects, the nature of adaptation learning, raises the issue as to whether control mechanisms changed as a result of the aiming constraints.

Consolidation processes were also implicated in two further studies where success perceptions were believed to have changed learning, but not performance. However, rather than there being benefits associated with practise to easier goals, disadvantages were shown for practice with difficult goals (compared to control conditions; Chiviacowsky & Harter, 2015; Chiviacowsky, Wulf & Lewthwaite, 2012). These findings were based on a coincident timing task, with success dependent on correct timing within 4 ms (difficult) versus 30 ms (easy). Although differences in retention may have been moderated by feelings of (lack of) success, supported by the fact that confidence was higher in the easier than difficult groups, it is also possible that a greater feedback dependence was created for the difficult group, in order to determine success within such a short time window, hindering intrinsic, error detection processes in the no-feedback, delayed tests of learning.

Despite the growing belief that enhanced perceptions of success change how people learn and what they remember, there has been reason to question the robustness of these effects, not only for reasons detailed above, but also based on difficulties in replication. It may be that these effects of success perceptions are very small or that they occur only under specific practice conditions and tasks. There have been published studies which have not shown any effects associated with manipulations to success and there are likely to be unpublished studies too, given the overrepresentation of positive effects in journal publications. In one published study where the “standard” results were not replicated, only participants’ awareness of the type of feedback they received, but not the content (three best or three worst trials), impacted learning (Patterson & Azizieh, 2012). These results contradicted those of Wulf and colleagues (Badami et al., 2012; Chiviacowsky et al., 2009; Chiviacowsky & Wulf, 2007; Saemi et al., 2012), where participants were also unaware of the content of their feedback. Carter and colleagues (Carter, Smith & Ste-Marie, 2016) also failed to show learning benefits associated with relatively better feedback. The absence of this effect was despite the fact that participants in the relatively better feedback group indicated higher judgments of learning than those in the relatively worse feedback group. Finally, in a study that involved manipulations to target size during dart-throwing, no differences in measures of accuracy were noted in performance or delayed tests of learning, despite a significant disparity in the success rates experienced by the large (82%) and small (32%) target groups during practice (Ong, Lohse & Hodges, 2015). Although the goal-criterion manipulation successfully changed the amount of success experienced and impacted self-efficacy during practice, these success perceptions did not impact learning.

One issue in this literature is the potential for participants to downplay the feedback or success manipulation due to competing error signals related to actual success and outcomes. For example, in a dart-throwing task, even if a target is large, the participant can still tell how close they are to the centre of the board which, regardless of target size, is still the best indicator of performance improvement. One of the few studies where outcome feedback is not readily available about success is in the stabiliometer, balance task (e.g., Lewthwaite & Wulf, 2010a;

Wulf, Chiviacowsky & Lewthwaite, 2012; Wulf, Lewthwaite & Hooyman, 2013). Although this task signals general balance performance (through proprioceptive feedback), average error over a trial period or time-on-target is difficult to ascertain. This is because trials last ~ 60 sec and measures of balance vary depending on the sensitivity of the feedback/measures (i.e., completely balanced (0 deg), or stable within 1 - 5 deg). Therefore, it may be that success-related effects on motor learning are moderated by task and the saliency of actual error feedback.

A second issue with manipulations about success, especially with respect to early differences that have emerged between groups (e.g., Chiviacowsky et al., 2009; Lewthwaite & Wulf, 2010a; Palmer et al., 2016; Saemi et al., 2012), is that rather than error (or accuracy) on a trial being the driving signal of success, it may be that improvement over trials best relates to perceptions of success (and learning). In tasks where goal-criterion or target size have been manipulated (e.g., Chiviacowsky et al., 2012; Chiviacowsky & Harter, 2015; Ong et al., 2015; Palmer et al., 2016; Trempe et al., 2012), participants assigned a difficult goal may not only experience low absolute scores, but also a lack of improvement. However, for targets that are only of moderate difficulty (e.g., Ong et al., 2015), there is more room for improvement, which might lead to positive perceptions and potentially wash out any negative effects associated with overall success or accuracy. Indeed, improving on a difficult task is likely to be more rewarding than not improving very much on an easy task. Therefore, there is a need in these studies to control and/or independently manipulate improvement in accuracy across trials.

In two studies, our aim was to test for both short and long-term effects related to enhanced success perceptions on motor learning through manipulations to: 1) accuracy feedback and the degree of improvement experienced during practice; and 2) feedback that encompassed comparative/peer results. We adopted a stabiliometer, balance task that had been used in one of the earliest studies because it showed both short and long-term benefits associated with perceptions of success (cf., Lewthwaite & Wulf, 2010a) and because participants are generally not privy to the actual outcome of their performance (such as overall time on target). In both studies, efforts were made to control for individual differences before practice manipulations through strict inclusion criteria. We restricted our recruitment to only females and non-competitive athletes with no prior training on balance tasks, in order to minimize between-subject variation as well as potential opposite-sex experimenter-participant interactions (e.g., Singer & Llewellyn, 1973; Stevenson & Allen, 1964).

Experiment 1

Our aim was to determine whether manipulations to accuracy feedback, based on the stringency of goal-criterion and degree of improvement across trials, would impact short- and/or longer-term learning and whether potential effects would correspond to changes in self-efficacy. In addition to Easy and Difficult goal-criterion groups, two groups practiced with the goal either increasing or decreasing in difficulty. In the former group, the degree of improvement would be low or absent, whereas for the decreasing group, marked improvements would be evidenced.

Measures of motivation and explicit knowledge were obtained to help distinguish any motivational versus reinvestment-related effects associated with the manipulations. Increased explicit knowledge in the Difficult group coupled with poorer performance, would point towards a motor-control related effect, consistent with reinvestment ideas, as would balance decrements under secondary task conditions for this group (associated with limited attentional capacity due to more conscious control of movement). A decrease in measures of motivation and self-efficacy (in the absence of increased explicit knowledge) would point towards a more social-cognitive effect (although these are not mutually exclusive). With the improvement-related manipulation, the decreasing difficulty group would experience a large improvement, such that greater perceptions of success towards the conclusion of practice should lead to enhanced skill consolidation and hence better balance performance in delayed retention. The reverse result is also possible if these groups were considered in terms of the reinvestment hypothesis. There is evidence that when practice schedules facilitate success or minimize error early versus later in practice, less explicit knowledge is gained and automaticity is encouraged (Maxwell et al., 2001; Poolton et al., 2005).

Methods

Participants and Groups

Females (*M* age of final sample = 22.6 yr, *SD* = 4.2; range = 18-34 yr, *N* = 48) without previous training in gymnastics or on balance devices and who were not competitive athletes responded to adverts. Participants had normal or corrected-to-normal vision and were not affected by current injuries or disorders that could affect balance. They were randomly assigned to one of four groups: Easy (goal criterion), Difficult, Easy-to-Difficult and Difficult-to-Easy groups. Six participants were removed from analyses; two were unable to participate on Day 2, one participant (Easy-to-Difficult), reported not believing the feedback (ascertained during debrief), and three participants were highly accurate on the first practice trial, showing error that was more than 2 SDs below the overall mean for the trial (n = 1 from Difficult, n = 2 from Difficult-to-Easy). These participants were replaced until there were 12/group (N = 48). Participants received $10/hour and informed consent was obtained in accordance with ethical procedures of the University's BREB.

Task and Apparatus

Participants stood on a stabiliometer (Lafayette Instrument, IN, USA) and tried to maintain it in horizontal during a 60 s trial. Feedback was provided in terms of % time-on-target (TOT). Unbeknownst to the groups, %TOT was based off either an easy goal (board within ±5° of horizontal), or a difficult criterion (within ±1° of horizontal)[2]. Two further groups received %TOT feedback according to sliding goal-criterion, either increasing in difficulty (from 5-1°; Easy-to-Difficult), or decreasing in difficulty (from 1-5°; Difficult-to-Easy) across 12 acquisition trials. Feedback was only given in practice.

Procedure

Acquisition. Participants started each trial with the platform in a horizontal position (sides off the ground), holding onto a safety bar located at waist level. Data collection started when participants released their hands from the safety bar and crossed them in front of their chest. Approximately two familiarization trials of 10 s were completed without feedback. Acquisition consisted of 12 trials of 60 s duration, interspersed with rest periods of ~60 s. After each trial, participants received verbal percentage time-on-target (%TOT) feedback, such as "you were balanced for 32% of the trial" and shown a monitor display of "TOT: 32 %". On a descriptor scale of "0" (for "not at all confident") to "100" ("very confident"; Zimmerman & Kitsantas, 1997), participants rated their confidence (i.e. self-efficacy) for achieving at least a certain %TOT score on the next trial, for a series of score iterations from 10 %TOT to the max of 100 %TOT, in increments of 10 (Bandura, 2006). These confidence probes were administered before trials 3, 6, 9 and 12. For a list of measures as a function of experimental phase, please see left of Table 1.

At the end of acquisition, participants filled out a 10-item, Likert scale (1 = "strongly disagree, 7 = "strongly agree") questionnaire consisting of the 8-item intrinsic motivation and amotivation subscales of the Situational Motivation Scale (SMS, Guay, Vallerand & Blanchard, 2000) and two customized items; whether they looked forward to practising a similar balance activity in future and whether they were motivated to do well during practice.

Delayed tests of learning. One day later, participants first performed two warm-up trials (10 s, no feedback) before completing five, 60 s retention trials without feedback. Before trials 1 and 4, participants provided confidence ratings of achieving %TOT scores in 10 % increments. Two secondary task trials followed, where participants performed an additional cognitive task of counting the number of high tones played in an audio sequence (consisting of random high or low tones per 1.5 s). They were instructed to perform both the primary task of balancing and the secondary counting task as well as they could, without trading off performance in one for the other. At the end of the experiment, participants were interviewed and asked to report any rules or techniques that they had generated/used during acquisition, and were questioned about the veracity of the feedback (see Table 1 for an overview of these measures).

Data Collection and Analysis

Platform deviations from zero were sampled with a potentiometer and rendered through LabVIEW software (version 9.0; National Instruments, TX, USA) at 1 KHz. From these raw data, root mean square error (RMSE), indicative of typical deviation from horizontal, and %TOT scores, that represented the duration of time the platform was within specific criterion angles of horizontal, were derived. Percentage TOT scores for 1, 3 and 5° were tabulated for all trials.

Separate repeated measures ANOVAs were applied with group as the between and trial as the within variable to analyze balance (RMSE, %TOT) and confidence measures in acquisition and retention. All participants ($n = 12$/group) were included in statistical analyses. Confidence on a trial was calculated by taking the average of all confidence ratings for the 10 score iterations (e.g., if 10% TOT was rated as 100 for "very confident", 20% TOT as 70 for "pretty confident" and 30% TOT as 0 for "not at all confident", the average rating would be 170/10 = 17%). A between-groups ANOVA was used to analyze mean balance measures for the secondary task trials. To assess for performance changes under secondary-task conditions, a repeated-measures ANOVA with group as the between and test (last retention trial, first secondary task trial) as the within variable was analyzed. Post hoc analyses were conducted using Tukey's HSD ($p <. 05$). Within variables were corrected with Greenhouse-Geisser for violations to sphericity. Partial eta squared (η^2_p) values are reported as measures of effect size.

To further assess whether the goal/feedback manipulation had impacted performance and learning, Pearson's correlational tests were conducted on the first and last acquisition trials and first acquisition and first retention trials. A high significant correlation would indicate a weak influence of the feedback manipulation and enduring influences of early individual differences. The correlation between first acquisition and first retention trial was calculated to give us a sense of the persisting effect of pre-existing balance capability on "learning". We assessed the first rather than the last retention trial, because participants improved across the retention trials (which may or may not have been related to the previous day's manipulation).

The accuracy of the tone-counting responses (number of secondary task trials with response errors), was assessed for between-group differences using the Kruskal Wallis test. This non-parametric test was also used to compare the four measures of motivation (intrinsic motivation, amotivation, motivation to do well, and motivation to engage in activity in future), as well as the number of rules. To assess for potential primary-secondary task trade-offs during the secondary task condition, point-biserial correlational tests were conducted on RMSE scores and corresponding tone-counting errors. A significant negative point-biserial correlation would reflect possible trade-offs between performance on the primary and secondary tasks. To aid readability of the results we have provided a table (Table 2) of all statistically significant effects.

Results

Success of the Feedback Manipulation

Feedback scores. Participants received different %TOT scores during acquisition that were congruent with our intended manipulation, as illustrated in Figure 1. Average %TOT feedback for the Easy group was higher (*M* = 54.6 %) than the Difficult group (*M* = 11.1 %) and the Difficult-to-Easy group showed a marked improvement in scores (from 6.8 to 63.2 %), whereas the Easy-to-Difficult group plateaued in scores around trials 4-5, after which scores declined. This was confirmed by main effects of group, trial, and a significant Group x Trial interaction (see Table 2). As can be seen from Figure 1 and confirmed through post-hoc testing, %TOT feedback for the Easy-to-Difficult group was not different from the Easy group in early practice, neither was %TOT feedback for the Difficult-to-Easy and Difficult groups. However, the changing-criteria groups started to show statistically significant deviations from the fixed criteria (Easy and Difficult) groups from the fifth trial onwards (ps<.05).

Self-efficacy. Confidence probes of self-efficacy closely mirrored %TOT feedback as illustrated in Figure 2. Again there were main effects of group, trial, and a Group x Trial interaction (see Table 2). Post-hoc analysis showed that the Easy and Difficult-to-Easy groups improved in confidence from T3 to T12. The Difficult and Easy-to-Difficult groups did not change in confidence from T3 to T12. The Easy and Difficult-to-Easy groups also showed a statistically significant increase in confidence from T6 to T12, while the Easy-to-Difficult group decreased in confidence. No significant change was noted for the Difficult group. Retention analysis yielded a similar pattern as observed at the end of acquisition (see right of Figure 2), again showing significant group differences.

No Evidence that Feedback Impacted Balance

All groups improved balance over acquisition and did not differ from one another (Figure 3). There was a significant trial effect but no significant effects involving group, Fs < 1. This was also true when we analyzed %TOT, based on various goal criteria (i.e., difficult = 1°, easy = within 5°, intermediate = 3°; see Table 2). The same was true for retention, where all group-related effects for both RMSE and other %TOT measures yielded Fs ≤ 1. Because there looked to be small, though non-significant, group differences for RMSE on T1 during acquisition, we ran additional ANCOVA analyses with T1 as a covariate. This did not change the group-related effects for RMSE or %TOT (all Fs < 1.1, ps > .39).

Performance on the secondary task also failed to differentiate the groups for RMSE and %TOT (Fs < 1). Participants unexpectedly improved in performance from the last retention trial (RMSE: *M* = 6.02, *SD* = 2.15) to the first secondary task trial (*M* = 5.34, *SD* = 1.95). Response accuracy for the secondary tone-counting task did distinguish the groups, $\chi^2(3, N = 48) = 8.50$, $p < .05$. It was the Easy group that made most errors (*M* = .83, *SD* = .58), although it was only significantly different to the Easy-to-Difficult group (*M* = .17, *SD* = .39). For comparison; Difficult (*M* = .42, *SD* = .52), Difficult-to-Easy (*M* = .58, *SD* = .67). There was a significant positive point-biserial correlation between RMSE on the first secondary task trial and accuracy of tone-counting response, $r_{pb} = .37$, $n = 48$, $p = .01$, but not on the second trial ($p = .58$). The direction of the relation indicates a tendency for high (or low) error on both tasks, rather than any trade-offs.

Participant's performance on the first practice trial before any feedback was received, continued to be correlated to final performance at the end of practice and in retention. Pearson's correlation analyses indicated large positive correlations between RMSE for the first and last practice trial, $r = .70$, $p < .001$, and first retention trial, $r = .75$, $p < .001$, accounting for ~50% of the variance in outcome measures.

Feedback Did Not Influence Motivation or Explicit Knowledge

The intrinsic (overall $M = 5.55$, $SD = 1.00$) and amotivation (overall $M = 2.18$, $SD = .97$) subscales of the SMS did not differentiate the groups (see Table 3 for group descriptive statistics), nor did the two custom items assessing motivation to do well (overall $M = 6.06$, $SD = 1.06$) and to practice in the future (overall $M = 5.73$, $SD = 1.11$). All indices showed that participants were motivated to do well. Rules and strategies reported (overall $M = 2.3$, SD = 1.0) also did not yield group differences ($p = .53$). Participants generally reported rules and strategies related to maintaining a low centre of gravity (e.g., bending of knees, slight forward lean), a wide base of support (e.g., turning feet out, feet about shoulder width apart), concentration (e.g., focusing on the fixation cross, not thinking "too much" about the task), and breathing (e.g., regular calm breaths, deep breaths to relax).

Discussion

Our aim was to determine if and how manipulations to feedback pertaining to goal-attainment influence success perceptions and learning of a stabiliometer balance task. Importantly, we assessed whether relative changes in feedback, reflecting more or lack of improvement over practice, would be more influential in determining success perceptions, and ultimately learning, than feedback just signaling absolute high or low error. The goal-criterion manipulation successfully influenced %Time-on-Target (%TOT) feedback and confidence.

Participants in the Easy and Difficult groups received the highest and lowest %TOT feedback scores during practice. The Difficult-to-Easy group experienced the largest relative improvement and scores gradually declined for the Easy-to-Difficult group. Confidence ratings mirrored these trends in practice and retention, even when feedback was not present.

There was no evidence that balance performance and learning were differentially impacted by these feedback manipulations and between group differences in confidence. There were no group differences in outcome measures for either retention or secondary task performance, or evidence of a negative correlation indicative of a potential trade-off between primary task performance (RMSE) and secondary task accuracy (errors). These findings contrast with studies where similar practice manipulations have produced immediate and/or delayed impacts on learning (cf., Chiviacowsky et al., 2012; Chiviacowsky & Harter, 2015; Palmer et al., 2016; Trempe et al., 2012). Yet, these data are congruent with other studies which have failed to show learning benefits associated with relatively increased perceptions of success (Carter et al., 2016; Ong et al., 2015, Patterson & Azizieh, 2012). Participants also did not report being more motivated as a result of these success-related manipulations, even though they believed the feedback to be valid. Because of the lack of group differences, we were unable to determine relative influences on performance due to reinvestment of knowledge/conscious processing (e.g., Masters, 1992) or success perceptions and positive affect (i.e., OPTIMAL theory; Wulf & Lewthwaite, 2016). It appears that performance was primarily a function of initial individual differences and attempts to correct based on error-nullification, regardless of the magnitude of the error or improvement across trials.

For the secondary task trials, there was a small, yet significant improvement in RMSE when the last retention trial was compared to the first secondary task trial, which was independent of group. Although we expected the secondary task of counting tones to destabilize balance, improvements in balance have also been associated with such secondary tasks (e.g., Fearing, 1925; Swan, Otani, Loubert, Sheffert & Dunbar, 2004). The addition of counting may have led to an increase in concentration (mental effort), which in turn positively benefited balance. Alternatively, the secondary task may have directed attentional focus away from the primary task, which facilitated balance due to the external focus (for a review see Park, Yi, Shin & Ryu, 2015). The Easy group also made the most tone-counting errors (although only significantly different to the Easy-to-Difficult group), which was not expected and goes against predictions related to reinvestment. Based on these data, we surmise that performance under secondary-task load conditions in this balance task might be ineffective in giving an indication of potential control strategies, related to reinvestment (cf., Masters, 1992; Maxwell & Masters, 2009). Because there were also no group differences in the amount of task-relevant rules and strategies generated to support performance, we do not have any evidence that the amount of error experienced or perceived in practice impacted knowledge accrual or attentional/movement control strategies.

One potential limitation of Experiment 1 was that performance was not evaluated against an explicit goal or standard. Though participants knew that higher %TOT meant better balance and that values close to 0% were poor and closer to 100% were good, and they could clearly see whether they were improving (at least in the Easy and Difficult-to-Easy groups), there was no specific criterion for interpreting "success". Although measures of confidence suggested that performance perceptions were impacted by the feedback, if we had asked participants how they thought they were doing (in comparison to peers) we may have been better able to determine how feedback (and hence performance) was interpreted. Therefore, in Experiment 2, false feedback of others' performance was provided as a standard of comparison and as a means to manipulate success perceptions.

Experiment 2

False social comparative feedback has been effective for bolstering success perceptions in other stabiliometer studies (Lewthwaite & Wulf, 2010a; Wulf et al., 2012; Wulf et al., 2013). In these studies, participants in a "positive" group received comparative feedback indicating they were performing better than others. This group showed immediate performance gains and sustained these benefits in retention, compared to a control group that received only veridical error feedback. Given other contradicting reports of a lack of behavioural effects with such success perceptions (Carter et al., 2016; Ong et al., 2015, Patterson & Azizieh, 2012), we first attempted to replicate the benefits of positive social comparative feedback in this task (see Lewthwaite & Wulf, 2010a).

In Experiment 2, we studied control processes related to platform adjustments (i.e., Mean Power Frequency, MPF) with higher frequency values thought to indicate greater automaticity (e.g., McNevin, Shea & Wulf, 2003). As with Exp. 1, balance was also assessed under secondary task conditions and participants were asked to verbalize any task-relevant knowledge at the end of practice. According to reinvestment-related explanations, compared to a no comparative feedback control group, we expected the Positive group to demonstrate higher MPF, less performance decrement under secondary task conditions, and less task-relevant knowledge.

After analyzing the data collected from the two groups, it was apparent that participants in the control group had perceived themselves to be as successful and motivated as those in the Positive group. Therefore we tested two further groups (Exp. 2b). We reasoned that by withholding all feedback, and providing only general negative comparative feedback, the two further groups (Negative feedback and a no-feedback control) would be differentiated on success perceptions and hence allow clearer conclusions about the effectiveness of this type of feedback manipulation. To better elucidate how success perceptions might be mediating any performance benefits we also collected heart rate data to give a psychophysiological measure of arousal. This measure allowed us to infer the effects of feedback on psychophysiology as well as relate any performance changes to increased arousal (in line with the OPTIMAL theory). Both positive and negative feedback were expected to be more arousing than their control conditions (as indexed by increased heart rate; Brehm & Self, 1989; Kahneman, 1973), with positive performance effects associated with increased motivation or effort expected in the positive group only.

Methods

Participants and Groups

Recruitment procedures were the same as Exp. 1. Only female participants (*M* age = 21.1 yr, *SD* = 3.4, range 18 – 33 yr) were recruited.

Exp. 2a: Participants were randomly assigned to either the Positive (RMSE + positive comparative feedback), or Positive (Pos)-Control group (RMSE only). One participant from the Positive group did not perceive the practice to be successful (neutral on Likert rating) and hence was removed from analysis and replaced. The final numbers were n=10/group.

Exp. 2b: After testing was completed for Exp. 2a, additional participants were randomly assigned to either a Negative (just negative comparative feedback) or Negative (Neg)-Control (no augmented feedback) group. Three participants from the Negative group perceived their practice to have been successful (above neutral Likert rating) and hence were not included in analysis and replaced. Two participants on the first acquisition trial were ~2 SDs above the group mean in the Negative group, and one participant was ~2SDs below the group mean in the Neg-

Control group. These individuals were also replaced such that the final numbers were n=10/group.

Task and Apparatus

The task and apparatus were identical to those in Exp. 1. In Table 1, an overview of groups and measures for Experiment 2 is provided.

Exp. 2a: For each acquisition trial, the Pos-Control group received error feedback, which represented typical angular deviation from horizontal (RMSE). The Positive group also received false, social comparative feedback about the “group average” of participants in our study on each acquisition trial. This closely mirrored the comparative feedback procedures of Lewthwaite and Wulf (2010a), although we applied a flexible rather than fixed (1.2) multiplier. Piloting revealed some issues in credibility with the latter procedure as a result of the exact increase and decrease in group error that mirrored the participant’s own performance, even when they got unexpectedly worse with practice. Hence, to ensure that the comparative feedback was credible and scaled to initial performance error, a flexible multiplier approach was applied to the first acquisition trial, followed by gradual decreases of ~ 1° (with variation) on subsequent acquisition trials. For initial errors between 9-16°, a multiplier of 1.5 was applied. Because the maximum angular deviation of the platform was 26.8°, using a smaller multiplier of 1.2 when initial error was high (RMSE = 16° or >) prevented the average error from exceeding the maximum. If participants performed exceptionally well on the first trial (RMSE = 9° or <), a larger multiplier of 1.8 was applied. This ensured that participants could perform better than the false average on every trial as they improved, and the participants understood that they had performed better than “others”.

Exp. 2b: Neither group received RMSE feedback. The Negative group was only given false comparative feedback that the trial was “below average performance”, with the exception of the first acquisition trial where they were told the trial was of “average performance”. This was designed to aid credibility and to alert to a relative lack of improvement compared to others. No feedback was provided to the Neg-Control group.

A photoplethysmography (PPG) sensor was attached to the distal phalanx of the middle finger for heart rate measures. Signals were wirelessly amplified and transmitted to a data acquisition module unit (BioNomadix MP150, Biopac systems Inc., QC, Canada) and recorded at a sampling rate of 1 KHz for further analyses (AcqKnowledge 4.2; Biopac systems Inc.).

Procedure

Acquisition. Procedures were the same as Exp. 1, except at the start of Day 1, participants were fitted with a PPG sensor, followed by one-minute of quiet standing to validate recordings. After initial familiarization trials, the acquisition phase consisted of seven 60 s trials to match the practice design of Lewthwaite and Wulf (2010a). However, as groups in the study by these authors were already differentiated in performance after one day of practice, which was sustained through the second day of practice and retention on Day 3, we eliminated the extra day of practice in our experiment. On conclusion of each trial, the RMSE score was displayed and read out by the experimenter in Exp. 2a. A second display of “group average” (RMSE) was provided and read to the Positive group only. For the Negative group (Exp. 2b), a general statement alerting the participant to the fact that their performance was below average in comparison to others who had previously completed the experiment was verbally provided after trials 2-7.

Before trials 2, 4 and 7, the Positive and Pos-Control groups were probed for their confidence in a similar fashion to Exp.1. They rated their confidence in achieving an RMSE score that was less than 20° on the next trial, followed by 18°, 16° etc. (in decrements of 2°). The

Negative and Neg-Control groups were asked for their expected comparative performance on the next trial in terms of whether it would be “at average”, “above average” or “below average” compared to other people who had taken part in the study.

At the end of acquisition, participants filled out a 7-point Likert scale questionnaire (1 = “strongly disagree, 7 = “strongly agree”) consisting of the perceived competency and task interest/enjoyment subscales of the Intrinsic Motivation Inventory (IMI; McAuley, Duncan & Tammen, 1989), the intrinsic motivation and amotivation subscales of the SMS (Guay et al., 2000), and the two customized items from Exp. 1. For a summary of measures, see Table 1.

Delayed tests of learning. After ~24 hours, participants completed the perceived competency subscale of the IMI before performing 7 retention trials without feedback. No warm up trial was provided on Day 2. Before retention trials 1, 4 and 7, the groups were asked the same confidence or expectation probes asked on Day 1. Following retention, one secondary task trial (identical to Exp. 1) was performed.

At the end of the experiment, participants were interviewed for any task-relevant rules, rated their outcome feedback (and/or group scores) with respect to credibility and provided a judgment of success pertaining to the end of practice on Day 1 (using Likert ratings from 1-7).

Data Collection and Analysis

Platform deviation was sampled and rendered in the same way as in Exp. 1. The balance outcome measures of mean RMSE and mean power frequency (MPF) were tabulated from the raw data. A rectangular window was applied to the fast fourier transform of the platform deviations to derive MPF. In all analyses, the comparative feedback groups were separately analyzed with their respective control groups, even though we illustrate all four groups in the figures that follow.

Separate repeated measures ANOVAs were applied to RMSE, MPF and confidence measures, in acquisition, retention and secondary task testing, with Group as the between variable, and Day and Trial, or Test as within variables. The non-parametric Mann-Whitney U test was used to analyze questionnaire data. Pearson's correlational tests were conducted on the first acquisition trial with both the last acquisition trial and first retention trial, to assess the impact of the feedback manipulation. Point-biserial correlational tests were conducted on RMSE and corresponding tone-counting errors to assess for potential trade-offs.

The entire HR waveform was low-pass filtered at 5 Hz and peaks were manually flagged across a 60 s trial to provide HR for each trial (bpm). As we were interested in the effect of the feedback manipulation on acquisition HR, the first acquisition trial on Day 1 served as the baseline/covariate in analyses. Psychophysiological data was missing[3] on 4.8 % of trials, either due to human error (e.g., forgetting to start collection), or equipment failure (e.g., detached electrodes). To aid with readability of the results, we have summarized statistically significant effects on the right side of Table 2.

Results

Success of the Feedback Manipulations

Self-efficacy and expectation. For Exp. 2a, self-efficacy/confidence probes showed a significant difference across trials, and testing days, as illustrated in Figure 4 (in both cases showing significant increases). However, the Positive feedback and its control did not differ in confidence ratings during acquisition or retention. There were no group-related effects, Fs<1.

In Exp. 2b, participants were asked whether they expected their next trial to be at, above or below average. We did not run statistical tests on these data, but present a descriptive analysis. Before T2, as illustrated in Figure 5A, all participants in the Negative group expected average performance on the following trial, but before T4, more than half expected below average performance and by T7, all participants expected below average performance. The trend for the Neg-Control group was somewhat opposite as expectations for average or above average performance increased from T2, to T4 and T7 (Figure 5B). These differences were maintained in retention without feedback, where ~70 % of people in the Negative group expected to be below average compared to less than 10 % in the Neg-Control.

Perceived competency. The Positive group (*M* = 6.02, *SD* = 0.56) perceived itself to be more competent in the balance task at the end of acquisition than the control (*M* = 4.85, *SD* = 0.78; see Table 2), but there were no group differences before retention. For the Negative groups, competency ratings were also collected before practice, but significant group differences were only shown when assessed at the end of practice and before retention. In both cases, the Negative group had lower ratings (Acq: *M* = 2.37, *SD* = .94; Ret: *M* = 2.80, *SD* = 0.90) than their control group (Acq: *M* = 4.32, *SD* = 1.20; Ret: *M* = 4.37, *SD* = 1.00).

Success rating and feedback credibility. There were no difference in mean perceptions of success measured at the end of testing between the Positive group (*M* = 6.50, *SD* = .53) and its control (*M* = 5.89, *SD* = .93). The Negative group did perceive lower success (*M* = 2.90, *SD* = .88) than their control (*M* = 4.50, *SD* = 1.35; see Table 2). Participants in all groups generally believed the feedback to be true (overall *M* = 5.39, *SD* = 1.21, *max* = 7). There were no group differences in credibility ratings.

Heart rate (HR). There were no group-related differences during quiet standing (T1 – baseline). During acquisition, the Negative group had higher HR than the Neg-Control (see Table 2 and Figure 6 for T1, covariate-adjusted means). No other group-related effects were significant for either comparative feedback group when compared to their respective control for acquisition.

No Evidence that Feedback Impacted Balance

Acquisition. As illustrated in Figure 7, RMSE decreased across acquisition trials for both experimental and control groups. However, in neither experiment were group-related effects observed.

There was also an increase in MPF (see Figure 8) for both experimental and control groups. Again, there were no group-related effects.

Delayed tests of learning. Participants continued to improve on RMSE over 7 retention trials on Day 2 as evidenced by a significant trial effect for the Positive and Pos-Control groups, and for the Negative and Neg-Control groups (see R1-R7, Figure 7), but again there were no group effects.

MPF showed a significant reduction over retention trials, for the Positive and Pos-Control groups, but not for the Negative groups, although there were no group-related effects.

The secondary task trial did not differentiate between either group and its control for RMSE and MPF, all *p*s > .28. There was a significant improvement in RMSE only for the Positive and Pos-Control groups for the secondary task trial ($M = 5.37$, $SD = 1.70$) compared to the last retention trial ($M = 6.00$, $SD = 1.90$). There were no differences in tone counting accuracy (*p*s > .52) and no correlation between RMSE and tone-counting errors for either experiment, *p*s > .35.

Variance of outcome measures explained by early practice performance.

Correlations between the first and last practice trials for all participants (Exp. 2a & b) were positive and relatively high for RMSE, $r = .52$, $p < .01$ and MPF, $r = .65$, $p < .001$

(accounting for ~27-42% of the variation in performance). Correlations between the first practice and first retention trials were similar; RMSE: $r = .42$, $p < .01$; MPF: $r = .54$, $p < .001$, underscoring the lack of influence of the feedback manipulations.

Negative (Not Positive) Feedback Impacted Motivation, But Not Explicit Knowledge

The Positive and Pos-Control groups did not differ on any of the motivation-related measures as presented in Table 3. The Negative group did show significantly lower intrinsic motivation compared to the Neg-Control, and significantly lower motivation to do well during practice (Negative: $M = 4.40$; $SD = 1.35$; Neg-Control: $M = 6.40$; $SD = .84$). There were no differences in mean interest/task enjoyment ratings or explicit knowledge, *p*s > .10 (see Table 3 for descriptives). The type of rules and strategies that were verbalized in Exp. 2 were similar to those detailed for Exp. 1.

Discussion

We first tested two groups on the balance task, with the aim of replicating the work of Lewthwaite and Wulf (2010a) and testing mechanisms underlying any potential benefits associated with enhanced success perceptions (OPTIMAL theory, Wulf & Lewthwaite, 2016). To maximize the chances of eliciting these benefits, which were absent in Exp. 1, we used positive social comparative feedback (Lewthwaite & Wulf, 2010a; Wulf et al., 2012; Wulf et al., 2013). Surprisingly, this feedback did not impact balance measures (RMSE and MPF) in immediate performance or learning. Neither self-efficacy nor success perceptions were enhanced, although there were transitory effects of positive comparative feedback on perceived balance competency. The control group in this study was as self-efficacious and motivated as the Positive group, arguably because receiving veridical RMSE feedback that signaled progress in performance was perceived as success. There was also no evidence that positive comparative feedback impacted explicit control processes, as there was no group difference in secondary task performance or explicit knowledge reported.

As a result of the good performance and perceptions by both groups in Exp. 2a, we anticipated that there might be a ceiling effect and hence we recruited two more groups and depressed success perceptions. Negative comparative feedback was given to one group (and no veridical feedback that would be indicative of improvement/success) for the control. Nevertheless, even when negative comparative feedback was effective in depressing outcome expectations, perceived competency, success and motivation, the Negative group and its control were not different in balance-related outcomes. These results are in line with recent findings indicating that enhanced perceptions of competency did not mediate any learning advantages associated with self-controlled practice, as would be expected based on the OPTIMAL theory (Carter & Ste-Marie, in press; Ste-Marie, Carter, Law, Vertes & Smith, 2016). In addition, motivation was not a significant predictor of motor learning in a study looking at the effects of (task irrelevant) choice in a throwing task (Grand, Daou, Lohse & Miller, in press).

As with Exp. 1, there was a small, yet significant improvement in balance performance for both Positive and Pos-Control groups when a secondary task was added, again casting doubts on the efficacy of our chosen secondary task to indicate attention-related control processes through dual-task interference for such balance tasks. There was also no indication of greater explicit control in the Negative group, as assessed through reported explicit strategies and rules accrued. As such, we had no evidence supporting the idea that negative feedback, or perceptions of (more) error changed control processes for this group in line with predictions from the conscious processing/reinvestment hypothesis (Masters, 1992; Masters & Poolton, 2012; Maxwell & Masters, 2009). In accordance with the OPTIMAL theory (Wulf & Lewthwaite, 2016), one of the processes through which positive feedback and enhanced expectancies should benefit motor learning would be through increased motivation, as potentially indexed by increased arousal (Brehm & Self, 1989; Kahneman, 1973). However, the positive group was not different to its control in either arousal or motivation. Only the negative feedback group showed greater arousal compared to its control, but this did not impact performance measures.

Although this Experiment was based on procedures adopted by Lewthwaite and Wulf (2010a), we did modify the procedures, which may have impacted the results. For example, we shortened the trials from 90 to 60 s (following trial durations used in Exp. 1) and gave only one day of practice. However, performance effects as a result of positive comparative feedback were present after just one day of practice in Lewthwaite and Wulf's study, which were maintained across the second practice session and retention. We also gave comparative feedback that was flexible rather than fixed to increase the validity of the feedback, which should have aided perceptions of success. We also limited our sample to only female participants, which should have minimized between-subject variability, facilitating statistical power. Individuals who deviated from the group average on initial performance and individuals who showed any skepticism in the credibility of the feedback were also removed to better control for individual differences. Thus, these relatively minor differences between the studies were unlikely to have been the reason for the disparities in the data.

In conclusion, these data from Experiment 2 raise serious doubts about the potential of feedback (positive, negative, general or comparative) to impact performance and learning in non-balance impaired populations in these tasks, even when success perceptions and motivation are moderated. Because of lasting impacts of initial performance differences (as indicated by correlation analysis), on later performance and learning, doubts are warranted about success-related feedback effects over and above any individual differences in initial performance.

General Discussion

Here we reported two sets of experiments where feedback manipulations had an impact on measures of self-efficacy (Exp. 1 & 2), success and intrinsic motivation (Exp. 2), but no impact on balance performance, learning, or motor control processes. In other stabiliometer balance studies where practice manipulations have led to changes in success perceptions and impacted learning, performance differences were noted early in acquisition (Lewthwaite & Wulf, 2010a; Wulf et al., 2013) or were close to significant in acquisition (Wulf et al., 2012). These performance-related effects are suggestive of potential individual differences in performance across the groups or effort-related benefits associated with enhanced success perceptions, whereby short-term improvements continue to impact performance in retention. Because our "more successful" groups did not show these immediate performance effects, based on these previous results, delayed learning benefits (as a result of enhanced consolidation due to perceptions of success) were unlikely. Therefore, at least in these balance tasks, feedback manipulations designed to impact success do not appear to impact motor learning through consolidation-related processes.

In two studies, where goal-criterion was manipulated, delayed learning benefits were elicited (Chiviacowsky et al., 2012; Trempe et al., 2012). In the study by Trempe and colleagues, participants adapted to a visuomotor rotation in the absence of visual feedback of their hand, completing consecutive aiming trials of short inter-trial durations (the next target appeared 1 s after participants moved their hand back to the start location). The task by Chiviacowsky and colleagues was a coincident-anticipation timing task, requiring participants to coincide the timing of a button press to the endpoint of a series of light stimuli without vision of their hand or the final trajectory of the light stimuli. Generally, participants who completed these laboratory tasks became dependent on the external feedback that they were provided. These tasks would not be ideal representations of real-world motor tasks that are inherently rich in intrinsic feedback.

Discrete aiming tasks have been used in other studies investigating success manipulations on learning. Even in these studies, the more competent or successful groups that were expected to benefit from the manipulation had demonstrated superior performance, or seemed to deviate in the direction predicted by the manipulation, as early as pre-test or after one block of practice (e.g., Chiviacowsky et al., 2009; Saemi et al., 2012; Palmer et al., 2016). This raises the question of whether these performance or learning benefits were a consequence of the experimental manipulation, or were merely manifestations of pre-existing group differences. For future research, or for examination of past data, we would argue that it is important to show that correlations between initial performance and later performance are low to moderate, as evidence that the manipulations are responsible for later emergent differences. In the current experiments, this correlational analysis yielded medium to high correlations, indicative of such persisting individual differences throughout practice, irrespective of feedback. Although this type of analysis is not without problems, where potentially more improvement is possible for people who start out with low accuracy (yielding low or negative correlations), high positive correlations, coupled with significant group effects, should of course be treated cautiously.

If success perceptions or expectancies moderated arousal and control processes related to the degree of "automaticity" in performance, we would have expected HR, MPF, secondary task performance and amount of accumulated explicit knowledge to distinguish our groups. In general, this was not the case. We had hoped that these measures would allow us to make conclusions about whether success perceptions were being mediated by processes associated with a conscious focus on correction of error (*cf*., implicit learning ideas, Masters, 1992; Masters & Poolton, 2012; Maxwell & Masters, 2009). However, because there were no differences in any of our outcome measures we were not able to distinguish between different theories concerning how success feedback may potentially function.

Although we did not see differences between a no-augmented feedback group and negative feedback group with respect to outcomes, performing badly, or receiving negative feedback can function in a positive fashion. There is evidence that self-doubt can motivate individuals to exert more effort after initial failure (Ede, Sullivan & Feltz, 2017) and that negative feedback can enhance motivation in individuals that lean towards learning/mastery goals (task achievement), rather than performance/ego goals (performance with respect to others; Ilgen & Davis, 2000). Conversely, enhanced success perceptions or better performance can lead to lower motivation or effort, as there is less incentive for learners to deploy the same amount of (or more) attentional resources or physical effort on a goal that is deemed easy to achieve (see review by Kluger & DeNisi, 1996). In stereotype threat research (Steele & Aronson, 1995), participants that are provided with negative conceptions about stereotypical ability on an experimental task associated with one's social group (e.g., gender, race) have been negatively impacted in their performance. While several accounts, such as explicit performance monitoring (Beilock, Jellison, Rydell, McConnell & Carr, 2006) and withdrawal of effort (Stone, 2002), have been proposed for this debilitating effect, there is also research showing that when participants are subjected to evaluation under stereotype threats, motivation is increased resulting in improved performance (e.g., Huber, Brown & Sternad, 2016; Jamieson & Harkins, 2007), provided that participants recognize the need to correct their dominant response and are given the time to do so (Harkins, 2006). Overall, the effect of negative feedback on motivation is complex, due to the influence of various contextual and personality factors. These factors need to be appreciated before methods become widely adopted promoting the unqualified use of success-enhancing feedback.

Overall, these data across three sub-sets of experiments, serve to cast doubts on the efficacy of success-related perceptions to impact motor performance and learning, especially when intrinsic feedback is available about performance. Manipulations of goal-criterion and social comparative feedback did influence self-efficacy and success perceptions but did not impact performance or learning. Moreover, measures of process did not point towards arousal (HR) or automaticity (MPF, explicit knowledge or secondary task performance) as candidate explanations for success-related changes in motor performance and control. The current findings add to other literature where there has been a failure to replicate the benefits of success or competency perceptions on motor learning (Carter et al., 2016; Ong et al., 2015; Patterson & Azizieh, 2012).

In terms of practical advice, even if success perceptions exerted little or no effect on learning, it still may be good advice for practitioners to adopt a positive coaching philosophy to encourage activity adherence, effort and persistency. What is potentially worrying about the advice that positive feedback about continual success and enhanced perceptions of competency with little effort should be encouraged is that this could lead to complacency (e.g., Beattie, Lief, Adamoulas & Oliver, 2011; Schmidt & DeShon, 2009) or disinterest from lack of challenge (Guadagnoli & Lee, 2004), which researchers have found to be detrimental to learning. Whilst good or positive feedback may affect perceptions of competency, this does not necessarily mean that more or less effort will be devoted to practice during acquisition, in that being better than others or improving a lot could lead to less effort than performing badly or not improving. This is further complicated by the fact that increased effort may be bad for performance (i.e., over-correcting for errors), especially in balance tasks where a more relaxed state associated with low conscious control may be preferable (see also ideas on implicit learning benefits; e.g., Maxwell & Masters, 2009). Practitioners should be cautioned against setting of unchallenging goals and avoid too much positive feedback indicative of success that is not a true reflection of performance.

References

Ávila, L. T., Chiviacowsky, S., Wulf, G., & Lewthwaite, R. (2012). Positive social-comparative feedback enhances motor learning in children. *Psychology of Sport and Exercise*, *13*(6), 849-853.

Badami, R., VaezMousavi, M., Wulf, G., & Namazizadeh, M. (2012). Feedback about more accurate versus less accurate trials: differential effects on self-confidence and activation. *Research Quarterly for Exercise and Sport*, *83*(2), 196-203.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. C. Urdan (Eds.), *Self-Efficacy Beliefs of Adolescents* (pp. 307-337). USA: IAP-Information Age Publishing, Inc.

Beattie, S., Lief, D., Adamoulas, M., & Oliver, E. (2011). Investigating the possible negative effects of self-efficacy upon golf putting performance. *Psychology of Sport and Exercise*, *12*(4), 434-441.

Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened?. *Personality and Social Psychology Bulletin*, *32*(8), 1059-1071.

Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature*, *382*(6588), 252-255.

Brehm, J. W., & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*, 109-131.

Cahill, L., & Alkire, M. T. (2003). Epinephrine enhancement of human memory consolidation: Interaction with arousal at encoding. *Neurobiology of Learning and Memory*, *79*(2), 194-198.

Carter, M. J., Smith, V., & Ste-Marie, D. M. (2016). Judgments of learning are significantly higher following feedback on relatively good versus relatively poor trials despite no actual learning differences. *Human Movement Science*, *45*, 63-70.

Carter, M. J., & Ste-Marie, D. M. (in press). Not all choices are created equal: Task-relevant choices enhance motor learning compared to task-irrelevant choices. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-017-1250-7.

Chiviacowsky, S., & Harter, N. M. (2015). Perceptions of competence and motor learning: performance criterion resulting in low success experience degrades learning. *Brazilian Journal of Motor Behavior*, *9*(1). 1-10.

Chiviacowsky, S., & Wulf, G. (2007). Feedback after good trials enhances learning. *Research Quarterly for Exercise and Sport*, *78*(1), 40-47.

Chiviacowsky, S., Wulf, G., Wally, R., & Borges, T. (2009). Knowledge of results after good trials enhances learning in older adults. *Research Quarterly for Exercise and Sport*, *80*(3), 663-668.

Chiviacowsky, S., Wulf, G., & Lewthwaite, R. (2012). Self-controlled learning: the importance of protecting perceptions of competence. *Frontiers in Psychology*, *3*, 1-8.

Criscimagna-Hemminger, S. E., Bastian, A. J., & Shadmehr, R. (2010). Size of error affects cerebellar contributions to motor learning. *Journal of Neurophysiology*, *103*(4), 2275-2284.

Ede, A., Sullivan, P. J., & Feltz, D. L. (2017). Self-doubt: Uncertainty as a motivating factor on effort in an exercise endurance task. *Psychology of Sport and Exercise*, *28*, 31-36.

Fearing, F. S. (1925). Factors influencing static equilibrium: An experimental study of the effect of controlled and uncontrolled attention upon sway. *Journal of Comparative Psychology, 5*(1), 1-24.

Grand, K. F., Daou, M., Lohse, K. R., & Miller, M. W. (in press). Investigating the Mechanisms Underlying the Effects of an Incidental Choice on Motor Learning. *Journal of Motor Learning & Development*. doi: 10.1123/jmld.2016-0041.

Gray, R. (2011). Links between attention, performance pressure, and movement in skilled motor action. *Current Directions in Psychological Science, 20*(5), 301-306.

Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior*, *36*(2), 212-224.

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*, *24*(3), 175-213.

Harkins, S. G. (2006). Mere effort as the mediator of the evaluation-performance relationship. *Journal of Personality and Social Psychology*, *91*(3), 436-455.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679-709.

Hosp, J. A., Pekanovic, A., Rioult-Pedotti, M. S., & Luft, A. R. (2011). Dopaminergic projections from midbrain to primary motor cortex mediate motor skill learning. *The Journal of Neuroscience, 31*(7), 2481-2487.

Huber, M. E., Brown, A. J., & Sternad, D. (2016). Girls can play ball: Stereotype threat reduces variability in a motor skill. *Acta Psychologica*, *169*, 79-87.

Ilgen, D., & Davis, C. (2000). Bearing bad news: Reactions to negative performance feedback. *Applied Psychology*, *49*(3), 550-565.

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of personality and social psychology*, *93*(4), 544-564.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*(2), 254-284.

Lewthwaite, R., & Wulf, G. (2010a). Social-comparative feedback affects motor skill learning. *The Quarterly Journal of Experimental Psychology*, *63*(4), 738-749.

Lewthwaite, R., & Wulf, G. (2010b). Grand challenge for movement science and sport psychology: embracing the social-cognitive–affective–motor nature of motor behavior. *Frontiers in psychology*, *1*: 42.

Lewthwaite, R., & Wulf, G. (2012). Motor learning through a motivational lens. In N. J. Hodges & A. M. Williams (Eds.), *Skill Acquisition in Sport: Research*, *Theory and Practice* (pp. 173-191). New York, NY: Routledge.

Masters, R. S. W. (1992). Knowledge, knerves and know-how: The role of explicit versus implicit knowledge in the breakdown of a complex motor skill under pressure. *British Journal of Psychology*, *83*(3), 343-358.

Masters, R. S. W., & Poolton, J. M. (2012). Advances in implicit motor learning. In N. J. Hodges & A. M. Williams (Eds.), *Skill Acquisition in Sport: Research*, *Theory and Practice* (pp. 59-76). New York, NY: Routledge.

Maxwell, J.P., & Masters, R. S. W. (2009). The explicit consequences of error in motor learning. In M. P. Simmons & L. A. Foster (Eds.), *Sport and Exercise Psychology Research Advances* (pp. 111-133). Hauppauge, NY: Nova Science Publishers, Inc.

Maxwell, J. P., Masters, R. S. W., & Eves, F. F. (2000). From novice to no know-how: A longitudinal study of implicit motor learning. *Journal of Sports Sciences*, *18*(2), 111-120.

Maxwell, J. P., Masters, R. S. W., Kerr, E., & Weedon, E. (2001). The implicit benefit of learning without errors. *The Quarterly Journal of Experimental Psychology: Section A*, *54*(4), 1049-1068.

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, *60*(1), 48-58.

McGaugh, J. L. (2000). Memory--a century of consolidation. *Science*, *287*, 248-251.

McGaugh, J. L., & Roozendaal, B. (2002). Role of adrenal stress hormones in forming lasting memories in the brain. *Current Opinion in Neurobiology*, *12*(2), 205-210.

McNevin, N. H., Shea, C. H., & Wulf, G. (2003). Increasing the distance of an external focus of attention enhances learning. *Psychological Research*, *67*(1), 22-29.

Murayama, K., & Kitagami, S. (2014). Consolidation power of extrinsic rewards: Reward cues enhance long-term memory for irrelevant past events. *Journal of Experimental Psychology: General*, *143*(1), 15-20.

Ong, N. T., Lohse, K. R., & Hodges, N. J. (2015). Manipulating target size influences perceptions of success when learning a dart-throwing skill but does not impact retention. *Frontiers in Psychology, 6*: 1378.

Palmer, K., Chiviacowsky, S., & Wulf, G. (2016). Enhanced expectancies facilitate golf putting. *Psychology of Sport and Exercise, 22*, 229-232.

Park, S. H., Yi, C. W., Shin, J. Y., & Ryu, Y. U. (2015). Effects of external focus of attention on balance: A short review. *Journal of Physical Therapy Science*, *27*(12), 3929-3931.

Patterson, J. T., & Azizieh, J. (2012). Knowing the good from the bad: Does being aware of KR content matter?. *Human Movement Science*, *31*(6), 1449-1458.

Poolton, J. M., Masters, R. S. W., & Maxwell, J. P. (2005). The relationship between initial errorless learning conditions and subsequent performance. *Human Movement Science*, *24*(3), 362-378.

Robertson, E. M., Press, D. Z., & Pascual-Leone, A. (2005). Off-line learning and the primary motor cortex. *The Journal of Neuroscience*, *25*(27), 6372-6378.

Saemi, E., Porter, J. M., Ghotbi-Varzaneh, A., Zarghami, M., & Maleki, F. (2012). Knowledge of results after relatively good trials enhances self-efficacy and motor learning. *Psychology of Sport and Exercise*, *13*(4), 378-382.

Schmidt, A. M., & DeShon, R. P. (2009). Prior performance and goal progress as moderators of the relationship between self-efficacy and performance. *Human Performance*, *22*(3), 191-203.

Shadmehr, R., & Holcomb, H. H. (1997). Neural correlates of motor memory consolidation. *Science*, *277*(5327), 821-825.

Singer, R. N., & Llewellyn, J. H. (1973). Effects of experimenter's gender on subject's performance. *Research Quarterly. American Association for Health, Physical Education and Recreation*, *44*(2), 185-191.

Ste-Marie, D. M., Carter, M. J., Law, B., Vertes, K., & Smith, V. (2016). Self-controlled learning benefits: exploring contributions of self-efficacy and intrinsic motivation via path analysis. *Journal of Sports Sciences*, *34*(17), 1650-1656.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797-811.

Stevenson, H. W., & Allen, S. (1964). Adult performance as a function of sex of experimenter and sex of subject. *Journal of Abnormal and Social Psychology*, *68*(2), 214-216.

Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self-handicapping in white athletes. *Personality and Social Psychology Bulletin*, *28*(12), 1667-1678.

Swan, L., Otani, H., Loubert, P. V., Sheffert, S. M., & Dunbar, G. L. (2004). Improving balance by performing a secondary cognitive task. *British Journal of Psychology*, *95*, 31-40.

Trempe, M., Sabourin, M., & Proteau, L. (2012). Success modulates consolidation of a visuomotor adaptation task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 52-60.

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews Neuroscience*, *5*, 483-494.

Wulf, G., Chiviacowsky, S., & Lewthwaite, R. (2010). Normative feedback effects on learning a timing task. *Research Quarterly for Exercise and Sport*, *81*(4), 425-431.

Wulf, G., Chiviacowsky, S., & Lewthwaite, R. (2012). Altering mindset can enhance motor learning in older adults. *Psychology and Aging*, *27*(1), 14-21.

Wulf, G., & Lewthwaite, R. (2016). Optimizing performance through intrinsic motivation and attention for learning: The OPTIMAL theory of motor learning. *Psychonomic Bulletin & Review*, *23*(5), 1382-1414.

Wulf, G., Lewthwaite, R., & Hooyman, A. (2013). Can ability conceptualizations alter the impact of social comparison in motor learning. *Journal of Motor Learning and Development*, *1*, 20-30.

Zimmerman, B. J., & Kitsantas, A. (1997). Developmental phases in self-regulation: Shifting from process goals to outcome goals. *Journal of Educational Psychology*, *89*(1), 29-36.

Footnotes

1: Hereafter we use the term “success perceptions” to encompass both competency and success-related perceptions.

2: To avoid a performance ceiling effect for the Easy group or too easy a criterion for the Difficult group, we piloted which criteria resulting in percentage TOT scores that mirrored the success rates of ~ 50% for successful groups and < 10 % for less successful groups reported in earlier studies (Chiviacowsky et al., 2012; Trempe et al., 2012).

3: Three participants (1 from Pos-Control and 2 from Negative group) were excluded from the general linear analysis (GLM) of repeated measures ANOVA due to missing heart rate data. A linear mixed model analysis, that retained the valid data from the excluded participants, showed the same effects as the GLM analysis.

Table 1. Overview of experimental groups, feedback and measures in Exp. 1 and 2 as a function of testing phase.

Variables	Exp. 1	Exp. 2
Groups (Feedback received):	1) Easy (%TOT) 2) Difficult (%TOT) 3) Easy-to-difficult (%TOT) 4) Difficult-to-easy (%TOT)	1) Positive (RMSE + Pos comparative) 2) Pos-Control (RMSE) 3) Negative (Neg comparative) 4) Neg-Control (None)
Measures		
Day 1: Pre-session	None	IMI (competency) for Neg groups
Acq	%TOT, RMSE Self-efficacy	RMSE, MPF Self-efficacy/Comparative expectations Heart rate
Post-session	SMS (intrinsic & amotivation)	SMS (intrinsic & amotivation) IMI (competency, enjoyment)
Day 2: Pre-session	None	IMI (competency)
Ret, Sec	%TOT, RMSE Self-efficacy Accuracy Tone-counting	RMSE, MPF Self-efficacy/Comparative expectations Accuracy tone-counting
Post-session	Explicit knowledge Feedback credibility check	Explicit knowledge Credibility and success check

Acq = Acquisition phase; Ret = Retention test; Sec = Secondary task test; TOT = Time on target; Pos = Positive; Neg = Negative; RMSE = Root Mean Square Error; MPF = Mean Power Frequency; SMS = Situational Motivation Scale; IMI = Intrinsic Motivation Inventory

Table 2. Summary table of statistically-significant effects across Exp. 1 and 2.

Measures and Effects:		Exp. 1 Test statistic	Effect size (η^2_p)	Exp. 2 (a and b) Test statistic	Effect size
Feedback scores:					
	Group	F(3, 44) = 53.85***	.79		
	Trial	F(3.8, 166.1) = 44.96***	.51		
	Group x Trial	F(11.3, 166.1) = 45.26***	.76		
Self-efficacy/confidence probes:					
Acq:	Group	F(3, 44) = 23.62***	.62		
	Trial	F(2.1, 92.7) = 54.22***	.55		
	Group x Trial	F(6.3, 92.7) = 23.96 ***	.62		
Ret/	Group	F(3, 44) = 15.86***	.52		
Acq&Ret:	Trial	F(1, 44) = 12.71***	.22	(Exp 2a) F(1.3, 23.3) = 80.73***	η^2_p = .82
	Group x Trial	F(3, 44) = 6.07**	.29		
	Day			(Exp 2a) F(1, 18) = 19.63***	η^2_p =.52
RMSE:					
Acq:	Trial	F(6.0, 263.8) = 92.21 ***	.68	Fs > 40.00 ***	η^2_p > .68
Ret:	Trial	F(4, 176) = 6.56 ***	.13	Fs > 9.43 ***	η^2_p > .33
Sec:	Test	F(1, 44) = 17.19 ***	.28	(Exp.2a) F(1, 18) = 7.81 *	η^2_p = .30
Competency (IMI):					
End Acq:	Group			(Exp.2a) *Z* = 3.46, *U* = 4.50 ***	*r* = .77
	Group			(Exp.2b) *Z* = 2.91, *U* = 11.50 **	*r* = .65
Ret:	Group			(Exp.2b) *Z* = 3.07, *U* = 9.50 ***	*r* = .69

Success rating:			
Group		(Exp.2b) Z = 2.60, U = 16.50 **	r = .58
Heart rate:			
Group		(Exp.2b) F(1, 15) = 5.23 *	η^2_p = .26
Mean Power Frequency:			
Acq:	Trial	Fs > 2.35 *	η^2_p > .11
Ret:	Trial	(Exp.2a) F(3.3, 59.3) = 3.19 *	η^2_p = .15
Intrinsic motivation (SMS):			
	Group	(Exp.2b) Z = 2.25, U = 20.50 *	r = .50
Motivation to do well in practice:			
	Group	(Exp.2b) Z = 3.12, U = 10.00 **	r = .70

Key: Acq = Acquisition phase; Ret = Retention test; Sec = Secondary task test; RMSE = Root Mean Square Error; SMS = Situational Motivation Scale; IMI = Intrinsic Motivation Inventory. * p< .05; ** p< .01; *** p< .001.

Table 3. Descriptive statistics pertaining to the intrinsic and amotivation subscales of the Situational Motivation Scale (SMS) and explicit knowledge measures (i.e., number of rule and strategies reported) for Exp. 1 and 2.

	Exp. 1 / Groups				**Exp. 2 / Groups**			
Measures:	Easy	Difficult	Easy-Diff	Diff-Easy	Positive	Pos-Ctrl	Negative	Neg-Ctrl
Intrinsic motivation:								
Mean	6.02	5.58	5.38	5.23	5.55	5.95	4.35	5.25
SD	.61	.79	.83	1.48	.93	.86	.80	.84
Amotivation:								
Mean	1.88	1.79	2.52	2.54	2.23	1.98	2.90	2.40
SD	.70	.56	1.01	1.27	.70	1.01	.58	.84
# Rules/strategies:								
Mean	2.3	2.4	2.0	2.5	2.1	2.0	1.8	1.6
SD	.9	1.0	.8	1.5	1.0	.7	.9	.7

Diff = Difficult; Pos = Positive; Neg = Negative; Ctrl = Control.

Figure Captions

Figure 1 (Exp. 1). Average %TOT feedback (and SE bars) as a function of group and acquisition trials (T). SE bars for the Difficult group were too small to be visible on this scale.

Figure 2 (Exp. 1). Average confidence ratings (0-100%) across probe trials in acquisition (T) and retention (R). Error bars represent SEs.

Figure 3 (Exp. 1). Average RMSE (and SE bars) over the course of 12 acquisition (T1-T12) trials, 5 retention (R1-R5) trials, and 2 secondary task (S1-S2) trials.

Figure 4 (Exp. 2a). Average confidence ratings (0-100) across probe trials in acquisition (T) and retention (R). Error bars represent SEs.

Figure 5 (Exp. 2b). Percentage of participants in A) Negative feedback group and B) Neg-Control groups who expected at, below or above average performance on the next trial in acquisition (T) and retention (R).

Figure 6 (Exp. 2a&b). Mean covariate-adjusted HR as a function of group and acquisition trials, T2-T7). Error bars represent SEs.

Figure 7 (Exp. 2a&b). Average RMSE as a function of group and experimental phase: acquisition (trials T1-T7), retention (trials R1-R7), secondary task (Sec). Error bars represent SEs.

Figure 8 (Exp. 2a&b). Mean Power Frequency (MPF) of platform adjustments as a function of group and experimental phase: Acquisition (Trials T1-T7), Retention (Trials R1-R7), Secondary Task (Sec). Error bars represent SEs

Fig 1

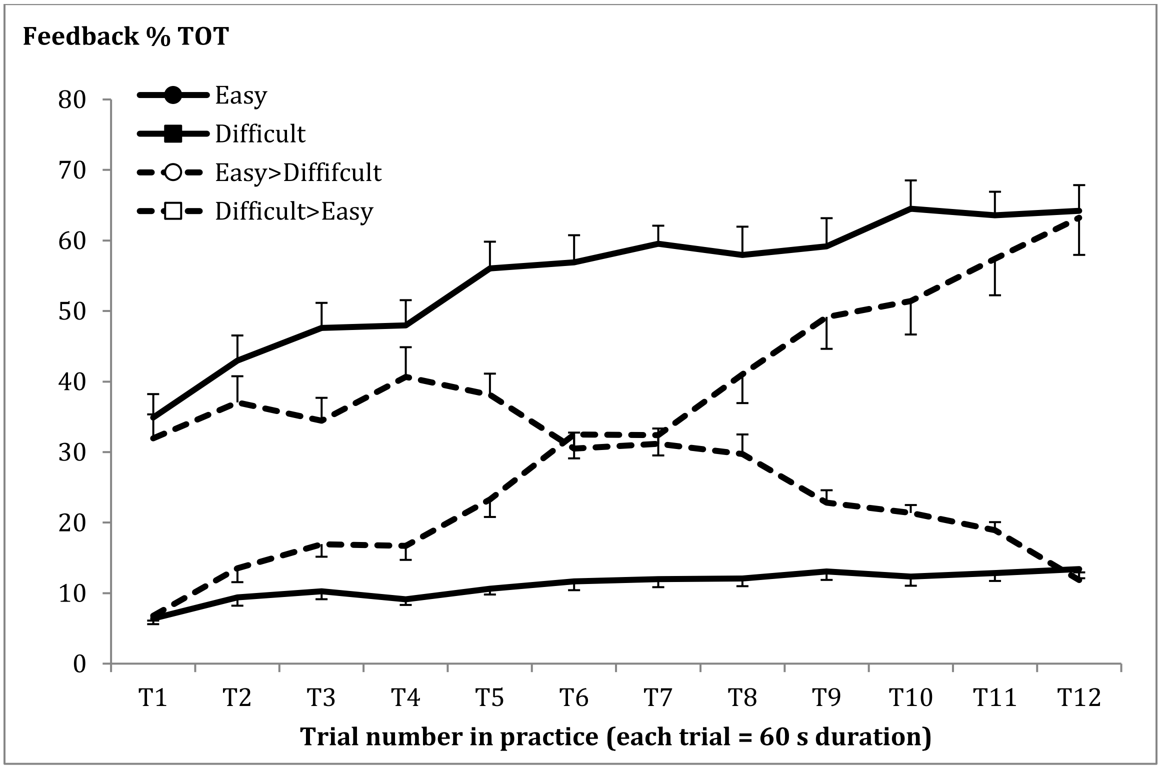

Feedback % TOT
80
70
60
50
40
30
20
10
0
Easy
Difficult
Easy>Diffifcult
Difficult>Easy
T1
T2
T3
T4
T5
T6
T7
T8
T9
T10
T11
T12
Trial number in practice (each trial = 60 s duration)

Fig 2

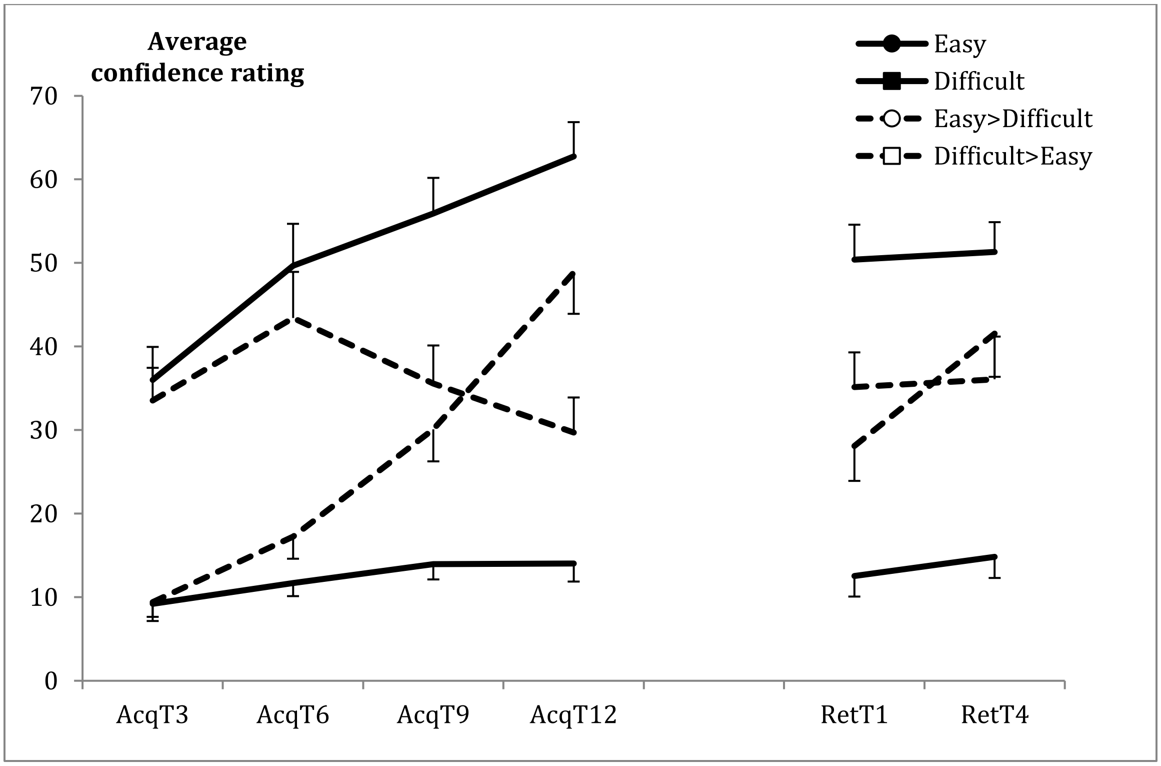

Average
confidence rating
Easy
Difficult
Easy>Difficult
Difficult>Easy
70
60
50
40
30
20
10
0
AcqT3
AcqT6
AcqT9
AcqT12
RetT1
RetT4

Fig 3

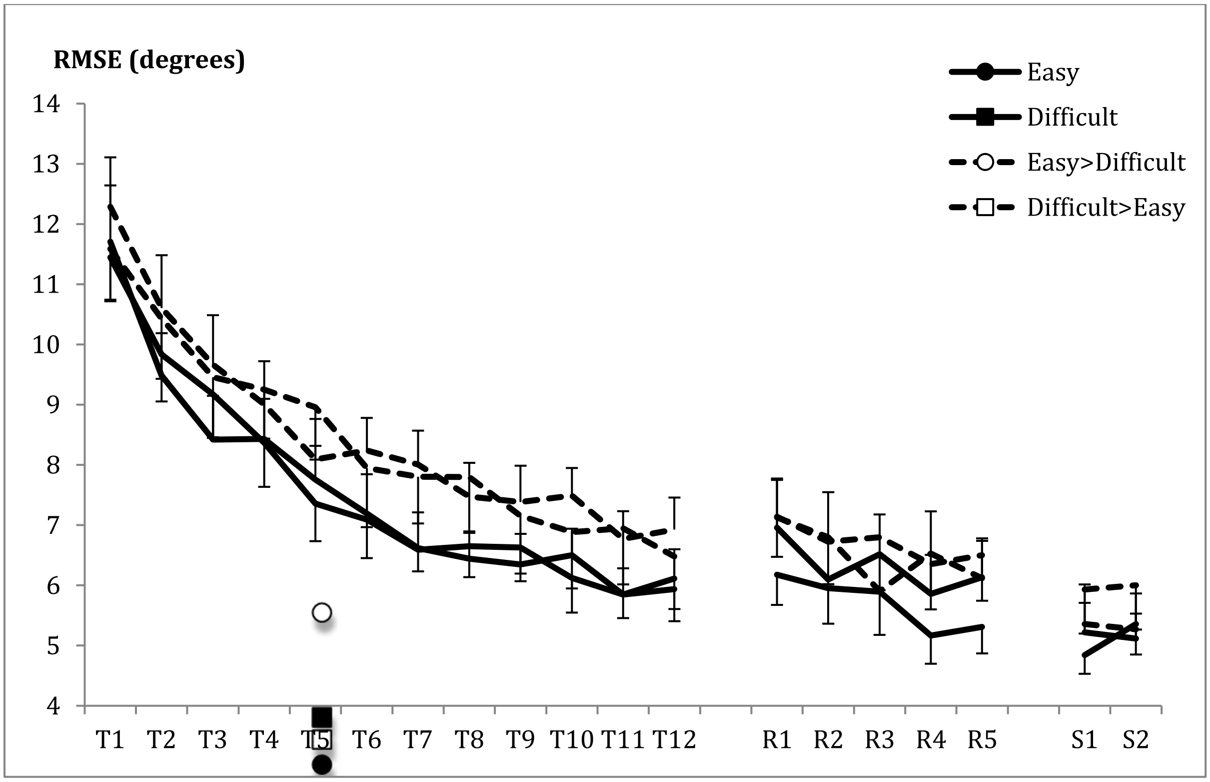

RMSE (degrees)
14
13
12
11
10
9
8
7
6
5
4
Easy
Difficult
Easy>Difficult
Difficult>Easy
T1 T2 T3 T4 T5 T6 T7 T8 T9 T10 T11 T12
R1 R2 R3 R4 R5
S1 S2

Fig 4

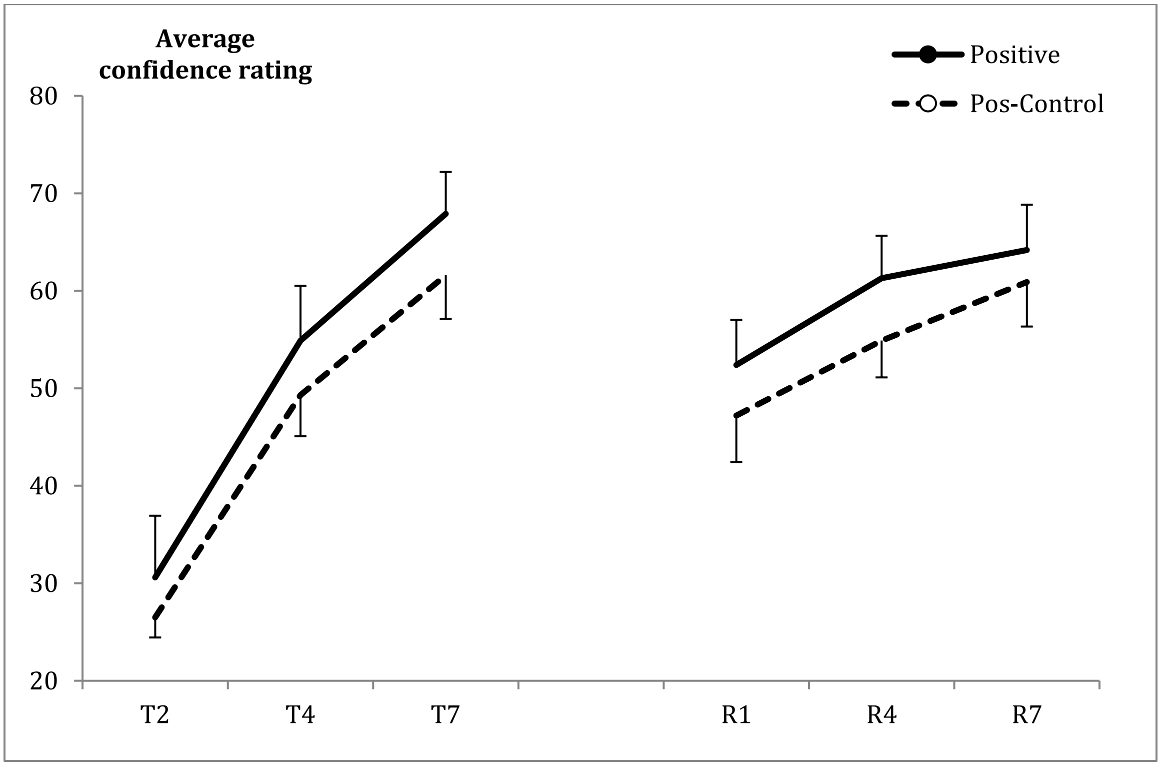

Average
confidence rating
Positive
Pos-Control
80
70
60
50
40
30
20
T2
T4
T7
R1
R4
R7

Fig 5a

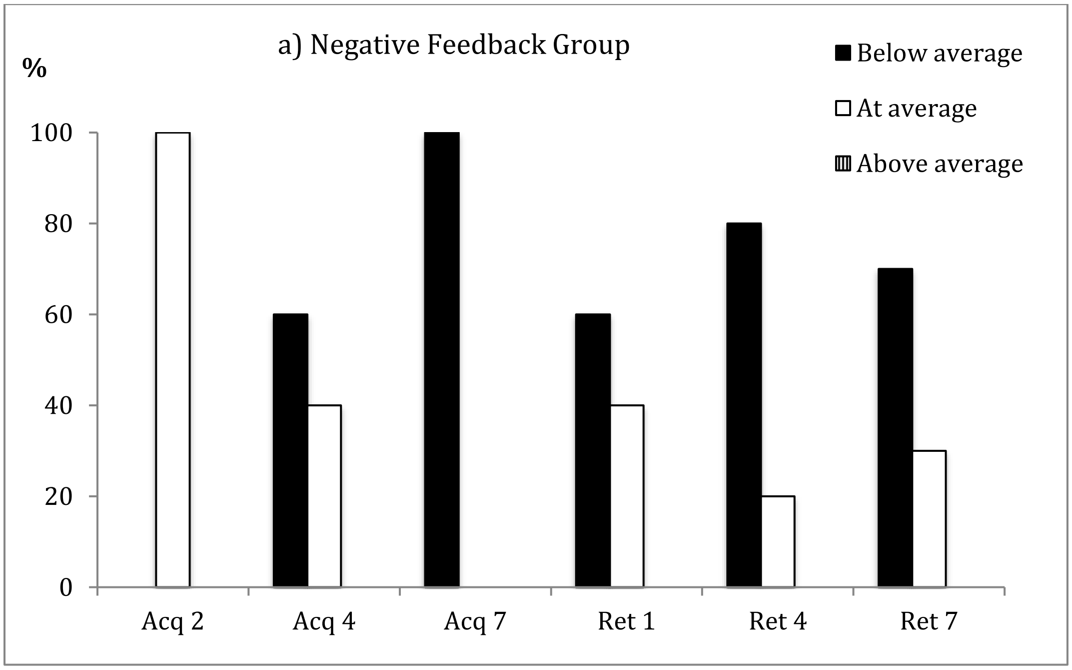

a) Negative Feedback Group
%
Below average
At average
Above average
100
80
60
40
20
0
Acq 2
Acq 4
Acq 7
Ret 1
Ret 4
Ret 7

Fig 5b

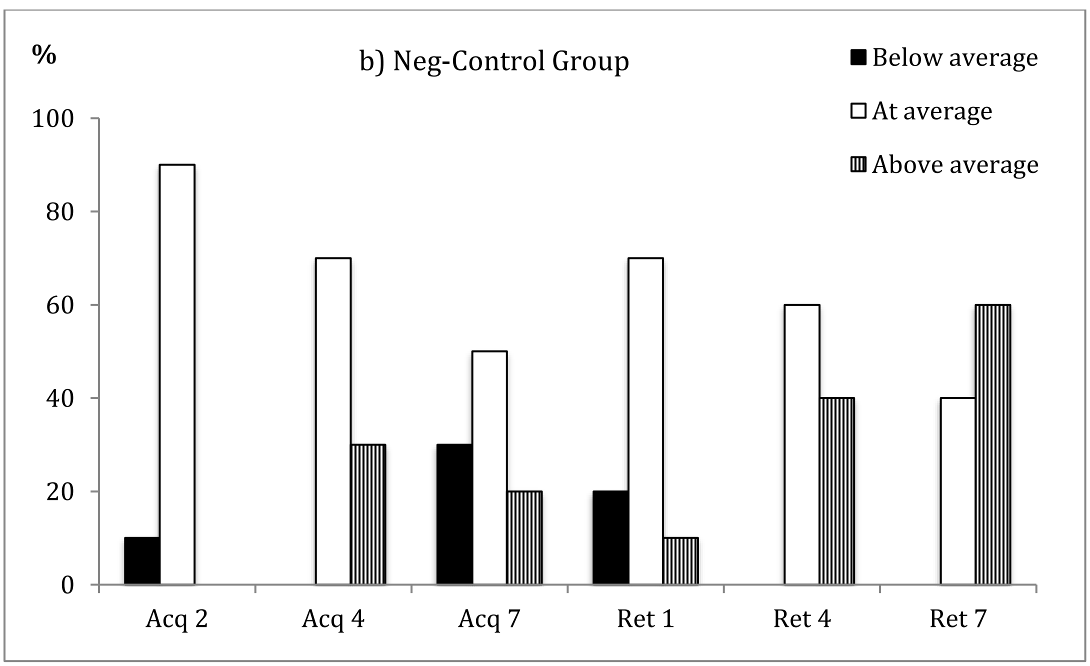

b) Neg-Control Group
%
Below average
At average
Above average
100
80
60
40
20
0
Acq 2
Acq 4
Acq 7
Ret 1
Ret 4
Ret 7

Fig 6

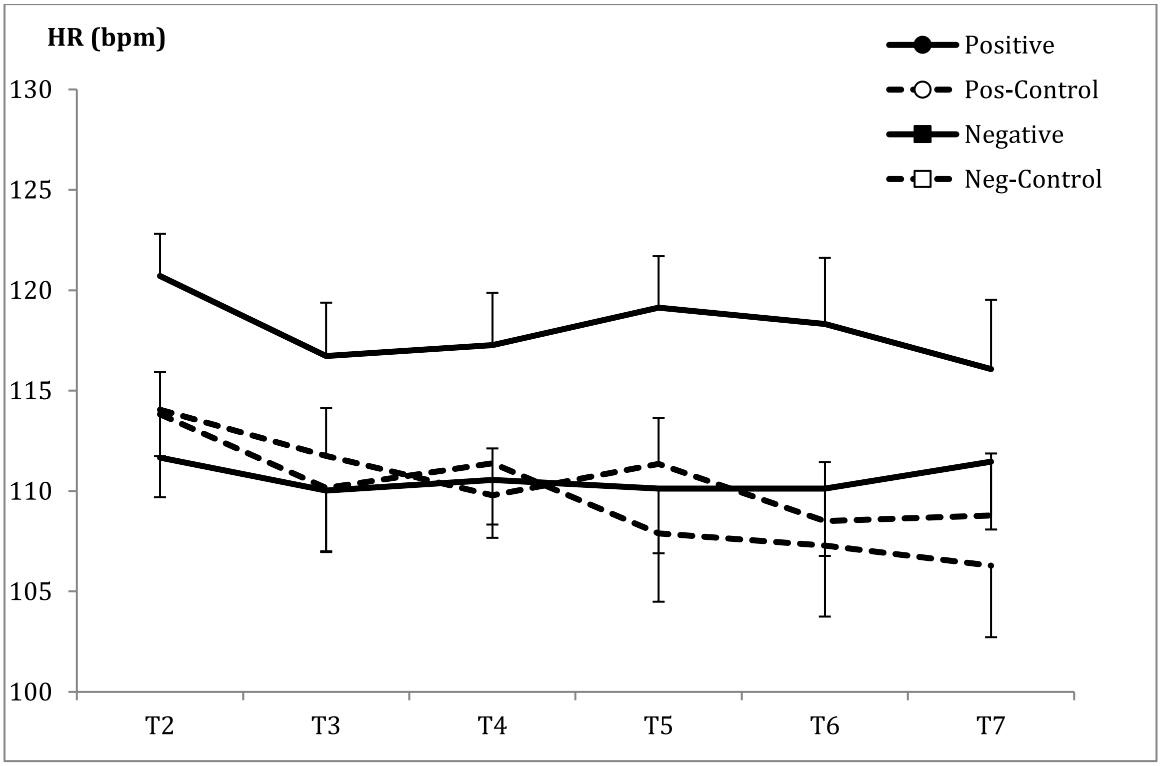

HR (bpm)
130
125
120
115
110
105
100
T2
T3
T4
T5
T6
T7
Positive
Pos-Control
Negative
Neg-Control

Fig 7

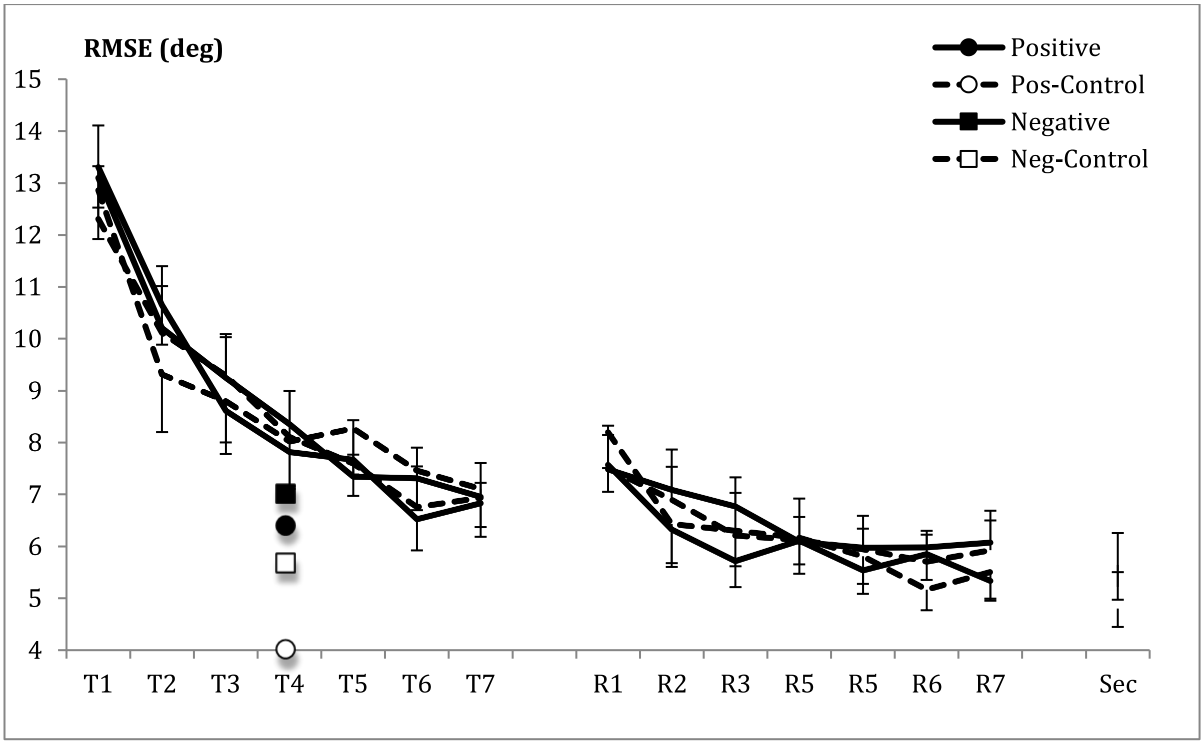
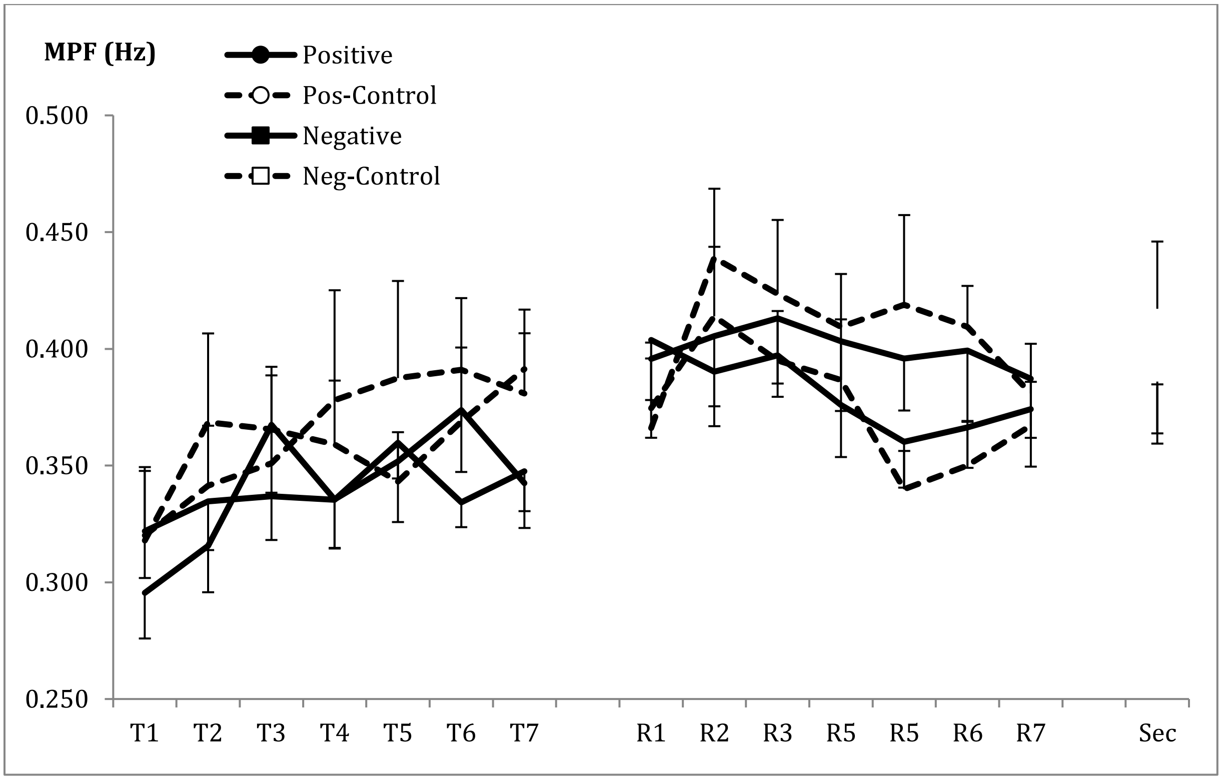

RMSE (deg)
15
14
13
12
11
10
9
8
7
6
5
4
Positive
Pos-Control
Negative
Neg-Control
T1 T2 T3 T4 T5 T6 T7
R1 R2 R3 R5 R5 R6 R7
Sec

Fig 8


MPF (Hz)
Positive
Pos-Control
Negative
Neg-Control
0.500
0.450
0.400
0.350
0.300
0.250
T1
T2
T3
T4
T5
T6
T7
R1
R2
R3
R5
R5
R6
R7
Sec